AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

High Prevalence and Genetic Diversity of Large phiCD211 (phiCDIF1296T)-Like Prophages in *Clostridioides difficile*

Julian R. Garneau,[a,c] Ognjen Sekulovic,[b] Bruno Dupuy,[c,d] Olga Soutourina,[c,d*] Marc Monot,[c,d] Louis-Charles Fortier[a]

[a]Department of Microbiology and Infectious Diseases, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

[b]Department of Molecular Biology and Microbiology, Sackler School of Medicine, Tufts University, Boston, Massachusetts, USA

[c]Laboratoire Pathogenèse des Bactéries Anaérobies, Institut Pasteur, Paris, France

[d]Université Paris Diderot, Sorbonne Paris Cité, Institut Pasteur, Paris, France

ABSTRACT *Clostridioides difficile* (formerly *Clostridium difficile*) is a pathogenic bacterium displaying great genetic diversity. A significant proportion of this diversity is due to the presence of integrated prophages. Here, we provide an in-depth analysis of phiCD211, also known as phiCDIF1296T, the largest phage identified in *C. difficile* so far, with a genome of 131 kbp. It shares morphological and genomic similarity with other large siphophages, like phage 949, infecting *Lactococcus lactis*, and phage c-st, infecting *Clostridium botulinum*. A PhageTerm analysis indicated the presence of 378-bp direct terminal repeats at the phiCD211 genome termini. Among striking features of phiCD211, the presence of several transposase and integrase genes suggests past recombination events with other mobile genetic elements. Several gene products potentially influence the bacterial lifestyle and fitness, including a putative AcrB/AcrD/AcrF multidrug resistance protein, an EzrA septation ring formation regulator, and a spore protease. We also identified a CRISPR locus and a *cas3* gene. We screened 2,584 *C. difficile* genomes available and detected 149 prophages sharing ≥80% nucleotide identity with phiCD211 (5% prevalence). Overall, phiCD211-like phages were detected in *C. difficile* strains corresponding to 21 different multilocus sequence type groups, showing their high prevalence. Comparative genomic analyses revealed the existence of several clusters of highly similar phiCD211-like phages. Of note, large chromosome inversions were observed in some members, as well as multiple gene insertions and module exchanges. This highlights the great plasticity and gene coding potential of the phiCD211/phiCDIF1296T genome. Our analyses also suggest active evolution involving recombination with other mobile genetic elements.

IMPORTANCE *Clostridioides difficile* is a clinically important pathogen representing a serious threat to human health. Our hypothesis is that genetic differences between strains caused by the presence of integrated prophages could explain the apparent differences observed in the virulence of different *C. difficile* strains. In this study, we provide a full characterization of phiCD211, also known as phiCDIF1296T, the largest phage known to infect *C. difficile* so far. Screening 2,584 *C. difficile* genomes revealed the presence of highly similar phiCD211-like phages in 5% of the strains analyzed, showing their high prevalence. Multiple-genome comparisons suggest that evolution of the phiCD211-like phage community is dynamic, and some members have acquired genes that could influence bacterial biology and fitness. Our study further supports the relevance of studying phages in *C. difficile* to better understand the epidemiology of this clinically important human pathogen.

KEYWORDS bacteriophage, *Clostridioides difficile*, *Clostridium difficile*, prophage genomics, phiCD211, phiCDIF1296T

Received 2 October 2017 **Accepted** 10 November 2017

Accepted manuscript posted online 17 November 2017

Citation Garneau JR, Sekulovic O, Dupuy B, Soutourina O, Monot M, Fortier L-C. 2018. High prevalence and genetic diversity of large phiCD211 (phiCDIF1296T)-like prophages in *Clostridioides difficile*. Appl Environ Microbiol 84:e02164-17. https://doi.org/10.1128/AEM.02164-17.

Editor Charles M. Dozois, INRS—Institut Armand-Frappier

Address correspondence to Louis-Charles Fortier, Louis-Charles.Fortier@USherbrooke.ca.

* Present address: Olga Soutourina, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, University Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette Cedex, France.

Bacteriophages (or phages) infecting the pathogenic bacterium *Clostridioides difficile* have received increasing attention lately due to their potential as alternative therapeutic agents and also because of their possible contribution to the biology and virulence of their host (1, 2). *C. difficile* still represents a major threat to human health, and the epidemiology of this pathogen is a topic of great clinical importance (3). Indeed, there is a lot of genetic and phenotypic diversity among isolates, and how such diversity affects virulence or clinical outcome is unclear (4). Prophages, i.e., integrated phages (viruses), contribute to the virulence of many bacterial pathogens and represent an important source of genetic diversity among strains of the same species (5). Previous studies in *C. difficile* suggest a great diversity of inducible prophages (6–9), some of which likely have some role to play in the biology of the host (10). For example, phages phiCD119 and phiCD38-2 were shown to repress and induce toxin production, respectively (11, 12). The phiCDHM1 and some related phage genomes encode genes that are predicted to participate in quorum sensing (13). The phiSemix9 phage genome was recently shown to carry a complete and functional binary toxin locus (CdtLoc) (14). There are currently 2,500+ complete or draft *C. difficile* genomes available in public repositories (15, 16). Analyzing all of them for the presence of prophages could provide very useful data to better understand the diversity and epidemiology of *C. difficile* and the potential contribution of phages to the biology and virulence of this pathogen.

A little more than 20 *C. difficile* phages have been fully characterized and their genomes sequenced (17–19). One general observation is that all *C. difficile* phages are temperate members of either the *Myoviridae* or the *Siphoviridae* family of the order *Caudovirales*, i.e., phages with contractile or noncontractile tails, respectively (17, 20), and most genomes are ~30 to 55 kbp in size. The *Myoviridae* phage genomes characterized so far generally share significant DNA homology and tend to form phylogenetically related clusters. On the contrary, a limited number of *Siphoviridae* phages have been described and sequenced (e.g., phiCD38-2, phiCD111, phiCD146, phiCD6356, and phiCD24-1 [9, 19, 21]), and they are more distantly related to each other genetically (22).

We had previously isolated and sequenced the genome of a large 131-kbp temperate phage that we called phiCD211 (accession no. NC_029048.2) (19). Almost concomitantly, Wittmann and colleagues reported a large phage genome of 131 kbp, called phiCDIF1296T (accession no. CP011970), that they identified as an episome while sequencing the genome of *C. difficile* strain DSM1296^T (also known as ATCC 9689) (23). The functionality of the phiCDIF1296T phage or the isolation and observation of phage particles was not reported by the authors, but it turned out that phiCDIF1296T and phiCD211 are the same phage and were found in the same original strain (DSM1296^T = ATCC 9689).

In the present study, we demonstrate that phiCD211/phiCDIF1296T is a functional phage and further confirm by electron microscopy observations of viral particles that it is a member of the *Siphoviridae* family of phages. We used PROKKA (24) with the curated PHASTER database (25) to improve the previous genome annotation. We also used the recently published PhageTerm software to determine the nature of the phage genome's termini (26), which were not available before. In addition, we used Bowtie2 (27) to screen raw sequencing reads from 2,584 *C. difficile* genomes available to determine the prevalence of phiCD211/phiCDIF1296T and related prophages. Finally, whole-prophage comparisons revealed the extreme genetic diversity and genome plasticity of phiCD211/phiCDIF1296T-like phages.

RESULTS

Prophage induction and phage isolation. Most *C. difficile* isolates harbor one or more prophages in their chromosome, either as integrated prophages or as episomal circular DNAs (6–9, 12, 17, 28, 29). We sought to determine whether strain ATCC 9689 (DSM 1296^T, here referred to as CD211) carried functional prophages. To investigate this, we used a classical mitomycin C induction strategy. Early-log-phase cultures (optical density at 600 nm [OD_{600}] of 0.1) of strain CD211 were treated with various

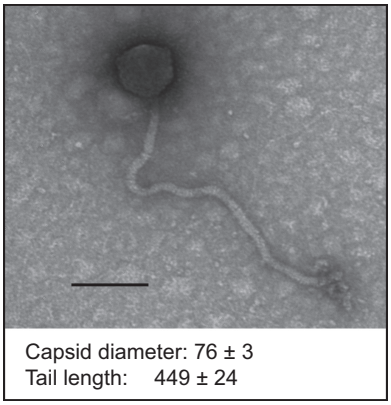


FIG 1 phiCD211 morphology as observed by TEM. Phage particles were stained with 2% uranyl acetate. The average sizes ± standard deviations of the capsid and tail were determined by measuring 5 different particles. The black bar represents 100 nm.

concentrations of mitomycin C, and we monitored the optical density over time. A typical lysis curve was observed at 1 μg/ml mitomycin C, suggesting the release of functional phage particles. We used this induction lysate in a spot test with multiple *C. difficile* isolates in search of a suitable replicating host. Unfortunately, despite screening over 50 different *C. difficile* isolates from various origins and ribotypes, none of them were susceptible to infection by the phage particles from the lysate.

Phage morphology. We analyzed the crude induction lysate under the electron microscope and found only one type of phage particle, which we named phiCD211, in reference to the host. The phage is composed of an isometric capsid with a diameter of 76 ± 3 nm and a long noncontractile and flexible tail of 449 ± 24 nm in length (Fig. 1). The morphological characteristics thus are reminiscent of the *Siphoviridae* family of phages of the order *Caudovirales* (30).

Complete genome sequencing. Because no suitable host was available to propagate phiCD211, we could not amplify and purify the phage from single plaques. However, since we observed only one type of phage particle in the induction lysate, we purified the phage DNA directly from the cleared lysate and completed whole-genome sequencing using the Illumina HiSeq platform at the Pasteur Institute in Paris. The sequence was deposited in the ENA public database on 9 December 2014 under the accession no. LN681537, and its sequence was later used and cited in our paper published in September 2015 (19). Of note, the complete genome sequence of phiCDIF1296T, a 131,326-bp phage identified in *C. difficile* strain DSM 1296^T, which is the same strain as ATCC 9689, from which we induced phiCD211, was published in August 2015 (23, 31). Not surprisingly, phiCDI1296T is 100% identical to phiCD211 at the nucleotide level over the entire genome. Although phiCD211 and phiCDIF1296T are identical, the phiCD211 genome was released in public databases prior to phiCDIF1296T, and it was cited almost simultaneously with phiCDIF1296T (19). Moreover, two following publications made reference to phiCD211 (22, 26). Therefore, although phiCD211 is identical to phiCD1296T, we kept the original phiCD211 name in this work to avoid confusion with previous publications.

Analysis of the phiCD211 genome. The double-stranded DNA genome of phiCD211 contains 131,326 bp and has a G+C content of 26.4%, which is slightly lower than the average G+C content of the host (~29%) (32). Aside from phiCDIF1296T, which is 100% identical to phiCD211, nucleotide BLAST homology searches did not retrieve other sequences with significant coverage, suggesting that phiCD211 is a unique phage in its group. As proposed by Rashid et al. (22), phiCD211 would be the representing member of a new phiCD211virus phage genus. Of note, the phiCDIF1296T phage genome has been detected as an episome in the DSM 1296^T strain in the course of a whole-bacterial-genome sequencing project (31). Therefore, and despite the presence of

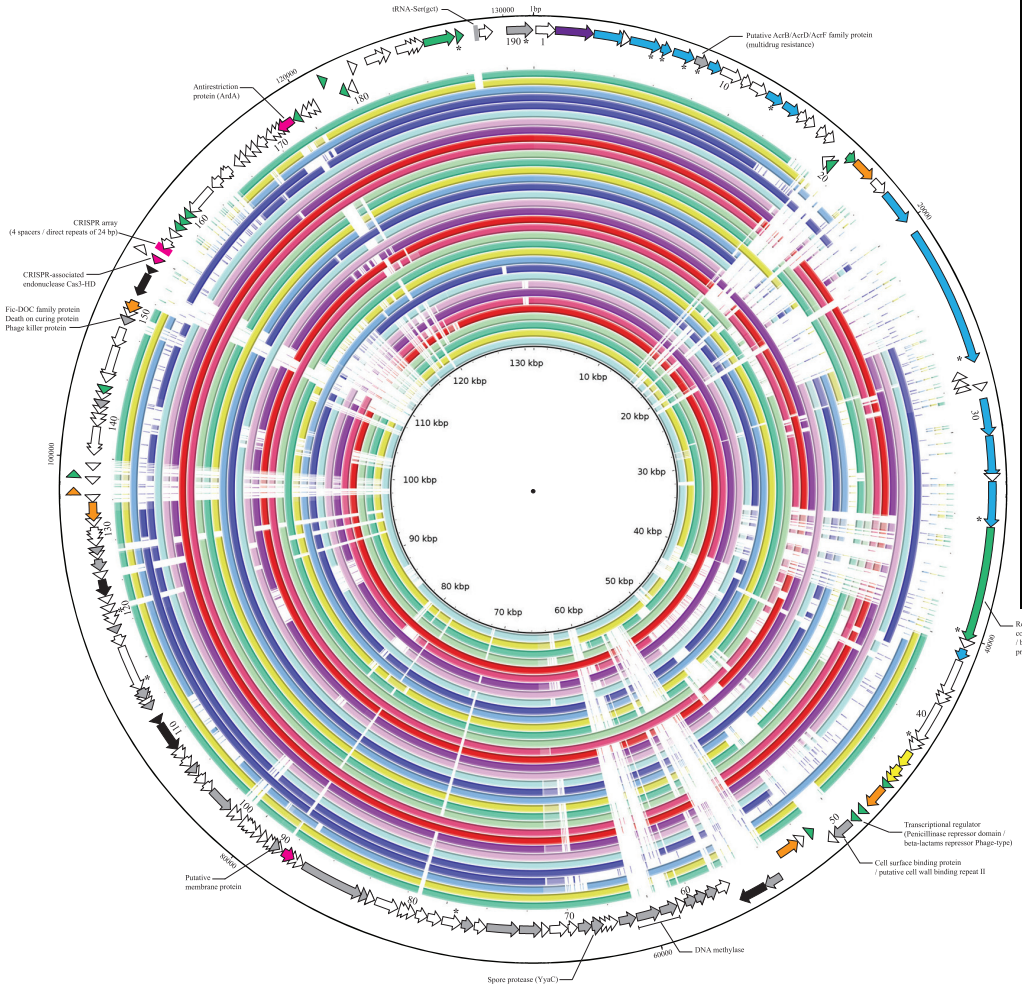
integrase genes, the data suggest that this phage does not integrate into the chromosome of its host (23). We have not determined experimentally whether phiCD211 was integrated or not in CD211, but since this strain is identical to strain DSM 1296^T, phiCD211 is likely episomal, like phiCDIF1296T.

Genome annotation. The phiCD211 genome encodes 190 putative open reading frames (ORFs), ranging in size from 34 amino acids (aa) to 2,176 aa. The few additional ORFs compared to the annotation of phiCDIF1296T (181 ORFs) (23) likely are due to differences in annotation algorithms and parameters. Of the 190 ORFs, 54 (28%) contained conserved domains according to InterProScan analyses, and a putative function could be assigned to 53 (28%) of them. Using PROKKA (24) and the PHASTER curated phage database (25), we obtained additional hits corresponding to proteins from known phages, even if distantly related to phiCD211. Overall, with this approach we could infer a function or find a related phage protein for nearly 40% of the predicted proteins. Detailed genome annotation is provided in Table S1 in the supplemental material, and a genomic map is shown in Fig. 2.

Annotation of the phiCDIF1296T genome has been detailed in the paper of Wittmann et al. (23), and here we will present only the features that were not previously described or for which further improvement can be provided. We identified 13 ORFs coding for structural proteins, and it is worth noting that ORFs 3, 5, 6, and 7 were related to the lactococcal phage 949, a large 115-kb *Siphoviridae* phage (33). Other ORFs were also related to large phages, like the structural protein ORF13 and the DNA polymerase III α subunit (ORF86), which were related to the 185-kb *Clostridium botulinum* phage c-st (34). Likewise, the hypothetical proteins ORF1, ORF161, and ORF189 were related to another large *C. botulinum* phage, D-1873 (35). Together, several gene products from phiCD211 matched to proteins found in other large phages, suggesting that structural features have been conserved among these large phages during evolution.

Among other interesting characteristics of the phiCD211 genome, 17 ORFs code for putative transcriptional regulators, suggesting that the regulation of the phage replication cycle is complex, and/or that the prophage could affect multiple bacterial functions. The phage also encodes a tRNA for serine (positions 128727 to 128817). Another striking feature of phiCD211 is the presence of 6 transposase genes, 3 of which (*orf56*, *orf152*, and *orf153*) are adjacent to 2 of the 3 integrase genes identified in the genome. The large ~55-kb genome fragment delimited by the transposase genes *orf56* and *orf153* is flanked by regions containing several transcriptional regulators, integrases, and lysis genes, as well as a putative CRISPR (clustered regularly interspaced short palindromic repeats)-associated endonuclease Cas3 protein and a nearby CRISPR array comprising 4 spacers of 42 or 44 bp. This configuration with several integrase genes and transposases suggests that this large genome fragment was acquired through one or more horizontal gene transfer events resulting from unprecise excision of a former integrated phiCD211 prophage, acquisition of transposons, or both. Further analysis of the CRISPR array revealed that spacer 1 targets known *C. difficile* phages, more specifically sequences from phages phiCDHM11, phiCDHM13, and phiCDHM14. The spacer is 100% identical to the target sequence, which is localized in the lysogeny module of these phages (17). Spacers 1 and 3 were also detected (100% identity) in several other *C. difficile* genomes, but spacers 2 and 4 did not match known sequences (Table S2). Of note, phiCDHM11, phiCDHM13, and phiCDHM14 were not detected in the strains carrying phiCD211, further supporting that the CRISPR array in phiCD211 is active and prevents further infection by these phages.

The phage genome also encodes 3 DNA methyltransferases (ORF62, ORF63, and ORF89) and an antirestriction protein (ORF173), which could participate in protection of the phage against host defense mechanisms. We also identified a death-on-curing protein (ORF149) of the Fic/DOC family (36) that provides further evidence for the episomal nature of phiCD211, and loss of the prophage should lead to death of the host. The PROKKA + PHASTER analysis also allowed the identification of a putative



Phage name	BRIG lane #	Gegenees Group ID
phiCD211	Reference Central black lane	17
phiCD211_like_ERR142363	1	6
phiCD211_like_ERR272205	2	7
phiCD211_like_ERR340289	3	8
phiCD211_like_ERR251746	4	9
phiCD211_like_ERR339557	5	10
phiCD211_like_ERR142205	6	11
phiCD211_like_ERR272067	7	12
phiCD211_like_ERR142055	8	13
phiCD211_like_ERR339813	9	14
phiCD211_like_ERR340201	10	15
phiCD211_like_ERR026385	11	16
phiCD211_like_ERR339803	12	5
phiCD211_like_ERR339682	13	18
phiCD211_like_ERR251727	14	19
phiCD211_like_ERR272061	15	20
phiCD211_like_ERR340162	16	21
phiCD211_like_ERR272222	17	22
phiCD211_like_ERR339785	18	23
phiCD211_like_ERR026357	19	24
phiCD211_like_ERR339506	20	25
phiCD211_like_ERR272197	21	26
phiCD211_like_ERR142074	22	27
phiCD211_like_ERR142261	23	28
phiCD211_like_ERR126258	24	29
phiCD211_like_ERR272134	25	30
phiCD211_like_ERR272126	26	31
phiCD211_like_ERR272085	27	32
phiCD211_like_ERR339796	28	33
phiCD211_like_ERR340254	29	34
phiCD211_like_ERR126241	30	35
phiCD211_like_ERR142303	31	1
phiCD211_like_ERR340169	32	2
phiCD211_like_ERR251779	33	3
phiCD211_like_ERR340061	34	4

FIG 2 Genome organization of phiCD211 and whole-genome comparison with other phiCD211-like phages. The outer tracks containing arrows represent the genomic map of phiCD211. The arrows indicate the predicted ORFs and their respective transcriptional orientation. ORF numbers are identified beneath arrow boxes in increments of ten. A color code was used for identification of predicted functions: purple, DNA packaging; blue, structural proteins; yellow, lysis; orange, lysogeny; green, transcriptional regulators and DNA binding proteins; black, transposases; pink, CRISPR-associated proteins and phage resistance; gray, proteins with other functions; white, putative proteins for which a predicted function could not be assigned and without conserved domains. Asterisks designate proteins that have been detected by mass spectrometry. BLAST Ring Image Generator (BRIG) was used to perform whole-genome comparisons. In the inner circles, the local similarity of each phage was calculated based on BLASTn high-scoring pairs and plotted against the reference phage (phiCD211), represented as the most central thin black line. Colors were assigned arbitrarily to the 34 phiCD211-like phages to help distinguish each of them. Track numbering goes from 1 (inner colored track) to 34 (outer colored track). The names of the phages along with their Gegenees group ID are listed in the table and include the identification of the bacterial strain in which they were detected.

protein of the AcrB/AcrD/AcrF family, which comprises efflux pumps generally involved in multidrug resistance (37). It is impossible at this point to comment on the impact of the prophage on antimicrobial susceptibility of *C. difficile*, since we have been unable to transfer the prophage into another host. Curing of the lysogen from its prophage could be attempted, but this approach would be quite challenging due to the presence of the Fic/DOC addiction system.

A putative YyaC spore protease (ORF68) was also identified in phiCD211. YyaC was shown to cleave small acid-soluble proteins (SASPs) in *Clostridium acetobutylicum* and acts like a germination protease (GPR) in *Bacillus subtilis* (38). Therefore, the presence of this protein in phiCD211 could affect sporulation and/or germination of the host. Interestingly, this strain sporulates poorly compared to other strains (39).

In summary, the phiCD211 genome possesses several interesting features, including several genes that could affect host physiology and fitness.

TABLE 1 Structural and virion-associated proteins detected by mass spectrometry

ORF	Function	No. of peptides[a]	Coverage (%)
5	Putative structural protein	3	6
6	Putative structural protein	10	73
7	Putative structural protein	12	32
8	Hypothetical protein	3	14
13	Hypothetical protein	7	31
25	Tail tape measure protein	5	3
33	Tail fiber protein	7	15
34	Hypothetical protein	4	3
41	Hypothetical protein	2	22
77	Hypothetical protein	2	6
116	Ribonucleoside triphosphate reductase	2	3
120	Hypothetical protein	2	13
188	Histone-like bacterial DNA-binding protein	2	10
190	Hypothetical protein	2	9

[a]Number of detected peptides that uniquely mapped to gene products from phiCD211. Only the proteins for which 2 or more peptides were detected are reported.

Structural protein analysis. We performed a mass spectrometry analysis in order to detect structural proteins composing the phiCD211 virion. We could not propagate the phage on a suitable bacterial host, so high phage titers were not available. Therefore, we concentrated the phage lysate on filter membranes and used this concentrated lysate for direct mass spectrometry analysis. A total of 14 proteins were detected, of which 5 were predicted to be structural proteins, and 3 others are possibly structural proteins as well (ORF8, ORF13, and ORF34) based on their genomic location (Table 1). A single peptide was also detected for ORFs 10, 12, 15, 31, 36, 37, and 40, which are also putative structural proteins based on their genomic location and/or bioinformatic predictions (not shown). However, only the protein hits for which at least two peptides were detected were reported in Table 1.

Analysis of genomic termini and packaging mechanism. New software designed to determine the nature of the physical DNA ends and phage genome packaging mechanism was developed. This software, called PhageTerm (26), was used to analyze the raw sequencing data obtained with phiCD211. The analysis revealed the presence of fixed left and right DNA ends with exact 378-bp direct terminal repeats (DTR) (Fig. S1). The DTR was not initially detected in phiCDIF1296T because the episomal prophage was sequenced (23), whereas phiCD211 was sequenced from the linear encapsidated viral genome, which contains the DTR.

Prevalence of phiCD211 in other *C. difficile* strains. We investigated the prevalence of phiCD211 in 2,584 published genomes from various *C. difficile* strains (19, 40). Table 2 reports the results obtained, along with the multilocus sequence type (MLST) inferred *in silico*. The phiCD211 and phiCD211-like prophages were detected in 5% of the strains analyzed (149 out of 2,584) and were distributed in 21 different MLSTs. Some MLSTs were more frequently lysogenized with phiCD211-like prophages, such as MLST 49 (13/29 strains, 45% prevalence), MLST 10 (14/43 strains, 33% prevalence), and MLST 58 (8/28 strains, 29% prevalence). This suggests that phiCD211-like prophages are widely distributed in *C. difficile*.

Whole-genome comparison of phiCD211-like phages. We next compared the 149 phiCD211-like phage genomes identified in *C. difficile* isolates to better appreciate their similarity. To do this, we first performed *de novo* assembly of the corresponding *C. difficile* genomes using short sequencing reads publicly available and then used the PHASTER detection tool to retrieve genomic regions containing the phiCD211-like phages (25). Using this approach, only 127 complete phiCD211-like genomes could be retrieved. The other 22 phiCD211-like phages detected by PHASTER were located on separate contigs that could not be joined due to assembly problems. Thus, they were not included for the whole-genome comparison because some parts could be missing or be misassembled. We next used Gegenees to compare the 128 genomes (including

TABLE 2 Prevalence of phiCD211 in 2,548 *C. difficile* isolates

MLST[a]	No. of occurrences of phiCD211[b]	No. of strains analyzed[c]	Prevalence[d] (%)
2	5	129	4
3	19	93	20
5	3	71	4
6	4	99	4
7	3	54	6
8	25	135	19
10	14	43	33
12	2	13	15
13	1	26	4
14	2	44	5
17	5	34	15
36	4	17	24
39	1	5	20
44	10	66	15
45	3	18	17
49	13	29	45
51	1	4	25
58	8	28	29
66	1	7	14
271	3	22	14
Unknown	22	201	11

[a]Multilocus sequence typing (MLST) has been inferred from raw sequences from 2,584 strains.
[b]Indicates the number of occurrences of phiCD211 in analyzed strains in each MLST group.
[c]Indicates the number of strains analyzed in each MLST group.
[d]Indicates the relative prevalence of phiCD211 in the corresponding MLST group.

phiCD211) at the nucleotide level (41). The three large phages 949 (*L. lactis*; GenBank accession no. HM029250.1), SPβ (*B. subtilis*; accession no. NC_001884.1), and c-st (*C. botulinum*; accession no. NC_007581.1) were included in the analysis. Gegenees uses an all-against-all BLASTn algorithm using short 200-bp fragments to assess whole-genome similarity and nucleotide identity level. The comparison matrix (heatmap) revealed the presence of distinctive clusters of $\geq$4 phages sharing $\geq$80% nucleotide identity over their entire genome (Fig. 3). Some of these clusters comprised several phages, like group 8, with 16 members. Other smaller groups contained only 2 or 3 members (e.g., groups 7, 9, 11, and 12). Phages that shared less than 80% identity in pairwise comparisons with other members were assigned a different number (e.g., groups 1, 2, 4, 5, etc.). Overall, 35 distinct phage groups were formed. The reference phage phiCD211 shared similarity with groups 16 and 18 but with slightly less than 80% with some members of these two groups, so we decided to assign it its own group (group 17) (Fig. 3). A table version of the Gegenees results as well as detailed information on the 35 phage groups are provided in Tables S3 and S4.

Representative members of the 35 distinct phage groups were also compared on a whole-genome scale using BRIG (Fig. 2) (42). This type of representation allows rapid identification of conserved and divergent regions in the various phage genomes. For example, we can see in Fig. 2 that the genes coding for the transposases (black arrows in the phiCD211 map) are not conserved in all phages. The genes encoding the DNA methylases (ORF62 and ORF63), the putative cell surface protein (ORF50), the ORFs encoding phage repressors possibly involved in lysogeny (ORF133 to ORF137), and the antirestriction protein (ORF173) all vary significantly between phiCD211 and the other phages. Note that because the reference phage in this analysis was phiCD211, genes that were present in other phages but that were absent from phiCD211 could not be visualized in the alignment. Therefore, additional genomic variations between the groups (insertions and deletions) likely exist. As an example, we identified one phiCD211-like phage with a significantly larger genome (phage 11, ERR339803, ~170-kb). When we used it as the reference genome in a BRIG analysis, other genomic differences were observed (Fig. S2).

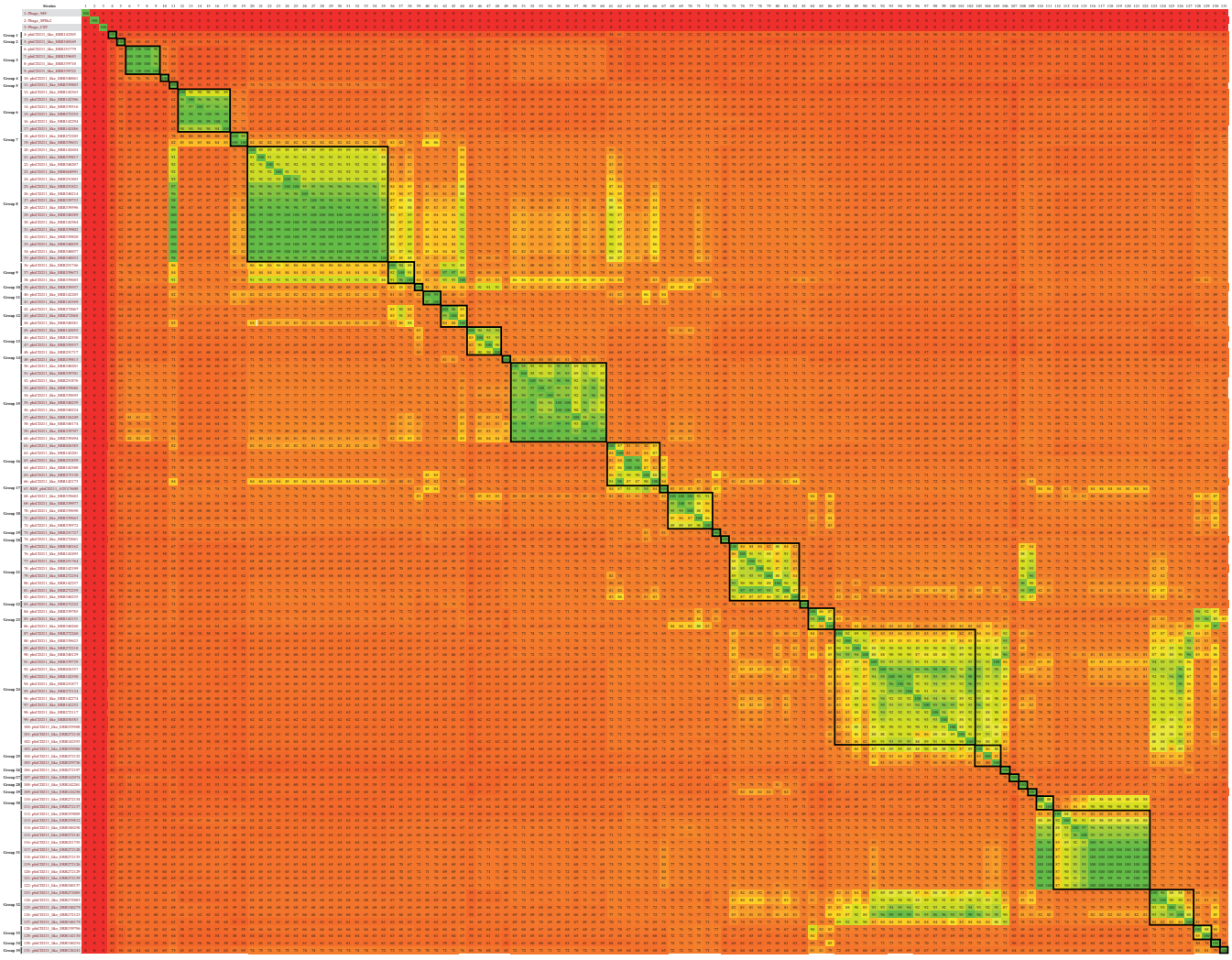


FIG 3 Whole-genome comparison using Gegenees. The heatmap shows the results for the comparative analysis of 127 phiCD211-like genomes. The reference phage phiCD211 and 3 genetically distinct and unrelated phages were also included in the comparison (total of 131 genomes). Similarity scores are based on a fragmented all-against-all pairwise fragment alignment using BLASTn and the accurate alignment option (fragment size, 200; step size, 100). The colors reflect this similarity, ranging from low (red) to high (green). The heatmap is asymmetric because genomes differ in sizes, and the similarity is calculated as a fraction of similar sequences in each genome. Highly similar phage genomes (bidirectional identity, ≥80%) are outlined with black lines. A cluster was defined when at least 4 phages shared an overall sequence identity of ≥90% over their entire genome, with a few exceptions at 88% (11 clusters were identified). Thirty-five distinct groups (including clusters of several phages) were also defined, and the corresponding group number is indicated on the left of a representative member. These groups are the same as those described in Fig. 2 and 4. The data used to build the heatmap are available in Excel format in Table S3 in the supplemental material.

We next selected one representative phage from the 11 groups that contained at least 4 members sharing ≥80% overall identity (Tables S3 and S4) and aligned them using Phamerator to get a more detailed view of the regions of homology (Fig. 4). The phiCD211 phage was included as a reference. This analysis enabled us to visualize regions of high similarity and to identify multiple insertions, deletions, and inversions between these groups of phages. The first 16 ORFs, among which some likely code for capsid proteins based on their genomic location (next to the terminase) (22), were highly conserved in all phage groups. Next to these ORFs, there was a region comprised of ORFs coding for tail structural proteins that seemed to diverge between the groups. The regions flanking ORF16 and ORF25, encoding the tail tape measure in phiCD211, seemed to be prone to recombination, which likely explains the variations observed in the intervening region. Likewise, ORF31, encoding a putative regulator of chromosome condensation, showed a high degree of sequence variability between the phage groups despite strong similarity in the flanking regions. This could also represent a hot

FIG 4 Whole-genome alignment of 12 representative phiCD211-like phages using Phamerator. Representative phage members from clusters comprising 4 phages or more (Gegenees groups [Fig. 3]) were included in the alignment. Amino acid comparisons of gene products were done with BLASTp and ClustalW functions encoded within Phamerator (62). Boxes of the same color denote proteins that belong to the same pham, and white boxes represent orphams (ORFs without assigned pham). A BLASTn pairwise alignment was used to detect regions of DNA similarity, and Phamerator used a spectrum of color shading to indicate the level of similarity, with purple showing the highest similarity and red showing the lowest similarity. Regions without color indicate the absence of DNA similarity using a BLASTn threshold limit of 10^{-4}.

spot of recombination that led to modular reorganization of phage genes. In agreement with this, a large genomic region has been inverted in phage ERR340162 (phage 75) compared to the other phages (Fig. 4). Also, several ORFs from phage ERR251779 (phage 6) were not conserved in the region comprising ORF19 to ORF80. This further supports the hypothesis that these regions represent hotspots of recombination which led to the exchange of functional modules, including structural genes composing the phage tail.

DISCUSSION

The morphology of phiCD211 had never been described before, although we had released the genome sequence of this phage in 2014 and cited the sequence in 2015 (19). In addition, the functionality of phiCD1296T had never been demonstrated, and only the genome sequence was published (23). With its large capsid measuring 76 nm in diameter and long tail of 449 nm in length, phiCD211 (phiCDIF1296T) represents the largest phage known to infect *C. difficile*, since all previously characterized phages have capsids with a diameter of ~50 nm and tails ~100 to 350 nm in length (6, 7, 9). Of interest, phiCD211 is morphologically similar to the *L. lactis* phage 949, the largest *Siphoviridae* phage known to infect this species, and has a capsid of 79 ± 7 nm in diameter and a tail of 500 ± 27 nm in length (33). Accordingly, some structural proteins from phiCD211 share sequence homology with structural proteins from phage 949 and other large clostridial phage (e.g., the *C. botulinum* phage c-st). Only a few other phages from different bacterial species have similar long tails, like the *Lactobacillus plantarum Siphoviridae* phage B2 that has a 500-nm tail (43), the *Bacillus cereus Myoviridae* phage 11 that has a 485-nm tail (44), or the *Bacillus megaterium Myoviridae* phage G, with a tail of 455 nm in length (45).

At 131 kbp, the phiCD211 genome is also the largest phage genome identified in *C. difficile*, although our PHASTER analyses revealed the presence of even larger phiCD211-like prophages in certain isolates (e.g., ERR339803 at ~170 kbp). However, the functionality of these larger phages needs to be determined experimentally. Of note, we could detect distant homology with certain phages only when using the PHASTER database comprising phage sequences only. BLASTp searches using larger nonredundant protein databases did not allow detection of some of these hits. Since phage genomes are often organized similarly (46), prediction of the function of certain distantly related proteins with limited homology is possible based on the genomic location of the corresponding hit, like structural proteins, for instance. Hence, for future phage genome annotation, we suggest using PHASTER to increase the strength of phage annotations. This would reduce the number of hits corresponding to proteins of unknown function.

Unfortunately, we were unable to find a suitable host to propagate phiCD211. This was not surprising considering the rather narrow host spectrum observed so far with *C. difficile* phages (9, 18, 47, 48). This may be due to the presence of wide-spectrum antiphage systems (49), including a functional CRISPR-Cas system (19, 50). It also likely depends on the presence of a suitable host receptor. It will be important at some point

to find a susceptible host to allow testing the influence of the prophage on bacterial phenotypes. Alternatively, curing the prophage from the host could be possible, but this might be more complex to do in *C. difficile* than in other bacteria, such as *Escherichia coli*, for which genetic tools are more accessible (51).

The genome of phiCD211 carries interesting genes that could influence drug resistance, like ORF8, encoding a putative AcrB/AcrD/AcrF protein. This family of proteins generally encodes multidrug resistance transporters, and their inactivation in *E. coli* was shown to affect biofilm formation and increase antimicrobial susceptibility (37). Also of interest, ORF46 encodes a putative EzrA septation ring formation regulator. During cell division and cytokinesis, polymerization of the cytoskeleton protein FtsZ is required to allow Z-ring formation that will eventually lead to proper septation at mid-cell. EzrA is conserved in low G+C Gram-positive bacteria and has been shown to bind to FtsZ to control its polymerization *in vitro* and *in vivo*. Overexpression of EzrA by 2-fold resulted in a filamentous cell phenotype in *B. subtilis* (52). Interestingly, the ATCC 9689 strain that we used in our study (CD211) forms filamentous cells in culture and sporulates very poorly (39), further supporting that the phiCD211-encoded EzrA homolog could be responsible for this phenotype.

A gene encoding a putative YyaC-like spore germination protease was also identified in phiCD211 (ORF68). Spore germination is an important step in the *C. difficile* life cycle, since it is required for vegetative growth, colonization, toxin production, and, thus, initiation of *C. difficile* infection (53). Small acid-soluble proteins (SASP) protect DNA within the spore and need to be removed during germination. In *Clostridium acetobutylicum*, a YyaC homolog has been shown to have proteolytic activity on SASP (38). Hence, ORF68 in phiCD211 possibly influences spore germination.

Finally, we identified a putative CRISPR-associated Cas3-HD endonuclease protein (ORF154) located close to two short CRISPR arrays. In type I CRISPR systems, such as in *E. coli*, the Cas3 nuclease acts in combination with the Cascade complex to nick, unwind, and degrade the target DNA (54). Cas3 homologs have been identified in the type I-B CRISPR system of *C. difficile*, but the function of this protein has not been demonstrated experimentally (19). It will be interesting to study the functionality of this phage-encoded Cas3 protein as well as the associated CRISPR arrays. CRISPR systems generally are encoded on the chromosome of the host, but some have been identified in phage genomes (55, 56). A number of *C. difficile* phage genomes carry CRISPR arrays (without *cas* genes), and here we present an example of a phage-encoded *cas* gene in *C. difficile* (19, 50). Whether it is active and plays some role in CRISPR-mediated phage interference will need to be investigated.

Whole-phage genomic comparisons revealed pervasive gene rearrangements in phiCD211-like prophages. Several gene insertions and deletions could be identified, as well as large DNA inversions. The presence of several putative transposases and integrases suggests that this large phage acquired additional genes through recombination with other phages and plasmids. Acquisition of large transposons that integrated into the prophage is also very likely. The presence of possible recombination hotspots probably further promoted genome evolution. This likely explains the chimeric nature and apparent genome plasticity of phiCD211-like prophages.

In conclusion, phiCD211 and phiCD211-like prophages are highly prevalent in *C. difficile* and might contribute to important phenotypes, including sporulation, germination, and antimicrobial susceptibility and resistance to invading DNA, such as phages. It will be very interesting to perform functional assays to validate some of the predictions we made using bioinformatic approaches.

MATERIALS AND METHODS

Bacterial strains and culture conditions. The *C. difficile* strain from which phiCD211 was isolated is ATCC 9689, but for simplicity and to be consistent with previous publications, we will refer to that strain as CD211 (9). Bacteria were routinely grown at 37°C in TY broth (3% tryptose, 2% yeast extract, pH 7.4) inside an anaerobic chamber (Coy Laboratories). All media were prereduced overnight under anaerobic atmosphere (5% CO_2, 10% H_2, and 85% N_2).

Prophage induction. To induce resident prophages from CD211, early-log-phase bacteria (OD_{600} of 0.1) were treated with 1, 3, and 5 μg/ml of mitomycin C (BioShop Canada), as previously described (9). The OD_{600} was monitored following mitomycin C addition to observe the typical drop in cell density that is characteristic of prophage induction (57). Crude lysates were filtered through 0.45-μm syringe filters (Sarstedt) and stored at 4°C.

TEM. Transmission electron microscopy (TEM) was performed as described before (9). Briefly, phage particles were washed 3 times in ammonium acetate buffer and transferred onto 400-mesh Formvar/carbon-coated copper grids. Negative staining was performed with 2% uranyl acetate, and phage particles were observed using a Hitachi H-7500 transmission electron microscope equipped with a Hamamatsu 10-megapixel digital camera controlled with AMT (Advanced Microscopy Techniques) software. The average size of viral capsids and tails was determined from five different images of isolated phage particles.

Phage DNA purification, genome sequencing, and annotation. Phage DNA was isolated from 50 ml of clarified induction lysate using the Qiagen Lambda minikit by following the manufacturer's recommendations. The quality of the DNA preparation (e.g., the absence of smears suggesting partial degradation) was verified by running an aliquot through a 0.8% agarose gel. The absence of contaminating bacterial genomic DNA was assessed by PCR amplification of the bacterial gene triosephosphate isomerase (*tpiA*) as described before (58). Single-end multiplex libraries (TruSeq DNA Library Prep kit) were created with a multiplexed protocol according to Illumina's specifications. The sequencing was performed on the Illumina HiSeq 2000 platform of the Pasteur Institute (Paris, France), which generated 110-bp reads. The sequencing reads were first scanned to remove the adaptor sequences and then were assembled *de novo* into a single contig using Velvet (59) (with the following options: Kmer, 107; -short; -fastq; -unused_reads; -read_trkg). We then used the MicroScope workflow (60) to generate an automatic functional annotation for each ORF, which was manually curated afterward. We first performed a standard annotation using BLASTp searches against the NCBI database and InterProScan conserved domain analyses. To get further insights on potential protein functions and relationships with other phages and to improve function assignment and overall annotation quality, we also performed a complementary genome annotation procedure using PROKKA (24) run locally with the most recent PHASTER-curated phage database (25) and using an E value threshold of 10^{-3}.

Analysis of structural proteins by mass spectrometry. Phage particles from 60 ml of crude phage induction lysate were concentrated to ~1 ml by ultracentrifugation for 1 h at 100,000 × *g* (Beckman SW 41ti rotor). The phage sample was further concentrated to a final volume of ~0.2 ml with Amicon Ultra 0.5-ml 3-kDa columns (EMD Millipore). Following concentration, the sample was digested with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) at the proteomic and mass spectrometry platform of the Cancer Research Pavilion at the Faculty of Medicine of the Université de Sherbrooke (Sherbrooke, Québec, Canada). Only the proteins for which at least two peptides were detected were kept in the final table.

Detection of phiCD211 in sequenced bacterial strains. The multilocus sequence types (MLST) (61) were inferred from raw sequences of 2,584 published strains (19, 40). To detect the presence of phiCD211 in other *C. difficile* strains, we mapped the raw sequencing reads of each strain on the sequence of the phage using Bowtie2 (27). A strain was considered to carry phiCD211 or a closely related prophage when the overall genome coverage was ≥78%. This value was determined after conducting preliminary analyses with different cutoff values. Generally, ≥80% gave robust results, but several additional phages could be detected in *C. difficile* strains if we slightly lowered the cutoff to 78% instead, so we decided to include them to expand the number of phages in our analyses.

Packaging mechanism and terminus analysis. The packaging mode and the nature of the phiCD211 genome termini were determined using PhageTerm software, version 1.0.8, with default options. PhageTerm uses randomly fragmented sequencing reads and the assembled phage genome to calculate coverage values in order to determine information on the termini (26).

Whole-phage genome comparisons and sequence alignments. Whole-genome nucleotide similarity analysis and generation of heatplots were performed using the Gegenees program, v2.2.1 (41) (with the following options: fragment size, 500; step size, 500; no identity threshold). The analysis was performed without threshold for the filtration of nonconserved sequences. Excluding this threshold means that the values shown in Fig. 3 are based on both genome size and genome conservation. A cluster was defined by at least 4 phage members with an overall pairwise sequence identity of ≥80%. This value was chosen based on our first screen using Bowtie2 that used a threshold coverage of 80%. Phage genomic maps and alignments were generated using the Phamerator program (62) with default options. Overall identity and E value were used to group similar gene products into phamilies (where pham indicates a family of homologous proteins with an identity threshold of ≥32.5% and an E value of $\leq 10^{-50}$). Phamerator also performed a pairwise DNA alignment between the genomes and used a spectrum of colors for shading areas to indicate the level of similarity, from purple for the highest similarity to red for the lowest similarity and no color when there was an absence of DNA similarity using BLASTn and a threshold value of 10^{-4}. Prior to performing the alignments, the phage genomes were reverse complemented when necessary and manually reorganized to set the start position at the terminase gene. Arbitrary numbers and colors were automatically assigned to each family (phams) by the program. Multiple-genome comparisons were also performed using BLAST Ring Image Generator (BRIG). The local similarity of each phage representative was calculated based on BLASTn high-scoring pairs and plotted against the reference phage, phiCD211 (42).

Accession number(s). The complete genome sequence of phiCD211 initially was deposited in the European Nucleotide Archive under accession no. LN681537 and in the National Center for Biotechnol-

ogy Information (NCBI) database under the accession no. NC_029048.1. This entry was later replaced with an updated version of the genome, starting at the physical genome termini determined with PhageTerm (25) and including the direct terminal repeats detected in the encapsidated viral particles (accession no. NC_029048.2).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02164-17.

SUPPLEMENTAL FILE 1, PDF file, 8.3 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.
SUPPLEMENTAL FILE 4, XLSX file, 0.2 MB.
SUPPLEMENTAL FILE 5, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was supported by a discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPIN-2015-06334 to L.-C.F.), the Institut Pasteur, the Agence Nationale de la Recherche ("CloSTARn," ANR-13-JSV3-0005-01 to O.S.), and the Institut Universitaire de France (to O.S.).

We thank Christiane Bouchier and Laurence Ma from the Genomics Platform and the Center of Informatics for Biology (Institut Pasteur).

L.-C.F. is a member of the Centre de Recherche du CHUS.

REFERENCES

1. Nale JY, Spencer J, Hargreaves KR, Buckley AM, Trzepiński P, Douce GR, Clokie MRJ. 2016. Bacteriophage combinations significantly reduce *Clostridium difficile* growth *in vitro* and proliferation *in vivo*. Antimicrob Agents Chemother 60:968–981. https://doi.org/10.1128/AAC.01774-15.
2. Sekulovic O, Fortier L-C. 2015. Global transcriptional response of *Clostridium difficile* carrying the ϕCD38 prophage. Appl Environ Microbiol 81:1364–1374. https://doi.org/10.1128/AEM.03656-14.
3. Evans CT, Safdar N. 2015. Current trends in the epidemiology and outcomes of *Clostridium difficile* infection. Clin Infect Dis 60(Suppl 2): S66–S71. https://doi.org/10.1093/cid/civ140.
4. Abou Chakra CN, McGeer A, Labbe A-C, Simor AE, Gold WL, Muller MP, Powis J, Katz K, Garneau JR, Fortier L-C, Pepin J, Cadarette SM, Valiquette L. 2015. Factors associated with complications of *Clostridium difficile* infection in a multicenter prospective cohort. Clin Infect Dis 61: 1781–1788. https://doi.org/10.1093/cid/civ749.
5. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H. 2003. Prophage genomics. Microbiol Mol Biol Rev 67:238–276. https://doi.org/10.1128/MMBR.67.2.238-276.2003.
6. Nale JY, Shan J, Hickenbotham PT, Fawley WN, Wilcox MH, Clokie MRJ. 2012. Diverse temperate bacteriophage carriage in *Clostridium difficile* 027 strains. PLoS One 7:e37263. https://doi.org/10.1371/journal.pone.0037263.
7. Shan J, Patel KV, Hickenbotham PT, Nale JY, Hargreaves KR, Clokie MRJ. 2012. Prophage carriage and diversity within clinically relevant strains of *Clostridium difficile*. Appl Environ Microbiol 78:6027–6034. https://doi.org/10.1128/AEM.01311-12.
8. Fortier L-C, Moineau S. 2007. Morphological and genetic diversity of temperate phages in *Clostridium difficile*. Appl Environ Microbiol 73: 7358–7366. https://doi.org/10.1128/AEM.00582-07.
9. Sekulovic O, Garneau JR, Néron A, Fortier L-C. 2014. Characterization of temperate phages infecting *Clostridium difficile* isolates of human and animal origins. Appl Environ Microbiol 80:2555–2563. https://doi.org/10.1128/AEM.00237-14.
10. Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence 4:354–365. https://doi.org/10.4161/viru.24498.
11. Govind R, Vediyappan G, Rolfe RD, Dupuy B, Fralick JA. 2009. Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. J Virol 83:12037–12045. https://doi.org/10.1128/JVI.01256-09.
12. Sekulovic O, Meessen-Pinard M, Fortier L-C. 2011. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. J Bacteriol 193:2726–2734. https://doi.org/10.1128/JB.00787-10.
13. Hargreaves KR, Kropinski AM, Clokie MRJ. 2014. What does the talking? Quorum sensing signalling genes discovered in a bacteriophage genome. PLoS One 9:e85131. https://doi.org/10.1371/journal.pone.0085131.
14. Riedel T, Wittmann J, Bunk B, Schober I, Spröer C, Gronow S, Overmann J. 2017. A *Clostridioides difficile* bacteriophage genome encodes functional binary toxin-associated genes. J Biotechnol 250:23–28. https://doi.org/10.1016/j.jbiotec.2017.02.017.
15. Didelot X, Eyre D, Cule M, Ip C, Ansari A, Griffiths D, Vaughan A, O'Connor L, Golubchik T, Batty E, Piazza P, Wilson D, Bowden R, Donnelly P, Dingle K, Wilcox M, Walker S, Crook D, Peto T, Harding R. 2012. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. Genome Biol 13:R118. https://doi.org/10.1186/gb-2012-13-12-r118.
16. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, Vaughan A, Golubchik T, Fawley WN, Wilcox MH, Peto TE, Walker AS, Riley TV, Crook DW, Didelot X. 2014. Evolutionary history of the *Clostridium difficile* pathogenicity locus. Genome Biol Evol 6:36–52. https://doi.org/10.1093/gbe/evt204.
17. Hargreaves KR, Clokie MRJ. 2015. A taxonomic review of *Clostridium difficile* phages and proposal of a novel genus, "Phimmp04likevirus." Viruses 7:2534–2541.
18. Hargreaves KR, Clokie MRJ. 2014. *Clostridium difficile* phages: still difficult? Front Microbiol 5:184. https://doi.org/10.3389/fmicb.2014.00184.
19. Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina Bedoya M, Fortier L-C, Severinov K, Dupuy B, Soutourina O. 2015. Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. mBio 6:e01112-15. https://doi.org/10.1128/mBio.01112-15.
20. Ackermann H-W. 2009. Phage classification and characterization. Methods Mol Biol 501:127–140. https://doi.org/10.1007/978-1-60327-164-6_13.
21. Horgan M, O'Sullivan O, Coffey Aidan Fitzgerald GF, van Sinderen D, McAuliffe O, Ross RP. 2010. Genome analysis of the *Clostridium difficile* phage PhiCD6356, a temperate phage of the *Siphoviridae* family. Gene 462:34–43. https://doi.org/10.1016/j.gene.2010.04.010.
22. Rashid SJ, Barylski J, Hargreaves KR, Millard AA, Vinner GK, Clokie MRJ. 2016. Two novel myoviruses from the north of Iraq reveal insights into *Clostridium difficile* phage diversity and biology. Viruses 8:310. https://doi.org/10.3390/v8110310.
23. Wittmann J, Riedel T, Bunk B, Spröer C, Gronow S, Overmann J. 2015. Complete genome sequence of the novel temperate *Clostridium difficile* phage phiCDIF1296T. Genome Announc 3:e00839-15. https://doi.org/10.1128/genomeA.00839-15.
24. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

25. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
26. Garneau JR, Depardieu F, Fortier L-C, Bikard D, Monot M. 2017. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep 7:8292. https://doi.org/10.1038/s41598-017-07910-5.
27. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
28. Hargreaves KR, Otieno JR, Thanki A, Blades MJ, Millard AD, Browne HP, Lawley TD, Clokie MRJ. 2015. As clear as mud? Determining the diversity and prevalence of prophages in the draft genomes of estuarine isolates of *Clostridium difficile*. Genome Biol Evol 7:1842–1855.
29. Hargreaves KR, Colvin HV, Patel KV, Clokie JJP, Clokie MRJ. 2013. Genetically diverse *Clostridium difficile* strains harboring abundant prophages in an estuarine environment. Appl Environ Microbiol 79:6236–6243. https://doi.org/10.1128/AEM.01849-13.
30. Ackermann H-W, Prangishvili D. 2012. Prokaryote viruses studied by electron microscopy. Arch Virol 157:1843–1849. https://doi.org/10.1007/s00705-012-1383-y.
31. Riedel T, Bunk B, Wittmann J, Thürmer A, Spröer C, Gronow S, Liesegang H, Daniel R, Overmann J. 2015. Complete genome sequence of the *Clostridium difficile* type strain DSM 1296T. Genome Announc 3:e01186-15. https://doi.org/10.1128/genomeA.01186-15.
32. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MTG, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet 38:779–786. https://doi.org/10.1038/ng1830.
33. Samson JE, Moineau S. 2010. Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. Appl Environ Microbiol 76:6843–6852. https://doi.org/10.1128/AEM.00796-10.
34. Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y, Ohnishi M, Ohtsubo E, Hattori M, Oguma K. 2005. The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. Proc Natl Acad Sci U S A 102:17472–17477. https://doi.org/10.1073/pnas.0505503102.
35. Fujii N, Oguma K, Yokosawa N, Kimura K, Tsuzuki K. 1988. Characterization of bacteriophage nucleic acids obtained from *Clostridium botulinum* types C and D. Appl Environ Microbiol 54:69–73.
36. Garcia-Pino A, Christensen-Dalsgaard M, Wyns L, Yarmolinsky M, Magnuson RD, Gerdes K, Loris R. 2008. Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. J Biol Chem 283:30821–30827. https://doi.org/10.1074/jbc.M805654200.
37. Bay DC, Stremick CA, Slipski CJ, Turner RJ. 2017. Secondary multidrug efflux pump mutants alter *Escherichia coli* biofilm growth in the presence of cationic antimicrobial compounds. Res Microbiol 168:208–221. https://doi.org/10.1016/j.resmic.2016.11.003.
38. Wetzel D, Fischer R-J. 2015. Small acid-soluble spore proteins of *Clostridium acetobutylicum* are able to protect DNA in vitro and are specifically cleaved by germination protease GPR and spore protease YyaC. Microbiology (Reading, Engl) 161:2098–2109. https://doi.org/10.1099/mic.0.000162.
39. Garneau JR, Valiquette L, Fortier L-C. 2014. Prevention of *Clostridium difficile* spore formation by sub-inhibitory concentrations of tigecycline and piperacillin/tazobactam. BMC Infect Dis 14:29. https://doi.org/10.1186/1471-2334-14-29.
40. Cairns MD, Preston MD, Hall CL, Gerding DN, Hawkey PM, Kato H, Kim H, Kuijper EJ, Lawley TD, Pituch H, Reid S, Kullin B, Riley TV, Solomon K, Tsai PJ, Weese JS, Stabler RA, Wren BW. 2017. Comparative genome analysis and global phylogeny of the toxin variant *Clostridium difficile* PCR ribotype 017 reveals the evolution of two independent sublineages. J Clin Microbiol 55:865–876. https://doi.org/10.1128/JCM.01296-16.
41. Agren J, Sundström A, Håfström T, Segerman B. 2012. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. PLoS One 7:e39107. https://doi.org/10.1371/journal.pone.0039107.
42. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402. https://doi.org/10.1186/1471-2164-12-402.
43. Nes IF, Brendehaug J, von Husby KO. 1988. Characterization of the bacteriophage B2 of *Lactobacillus plantarum* ATCC. 8014. Biochimie 70:423–427. https://doi.org/10.1016/0300-9084(88)90216-7.
44. Ahmed R, Sankar-Mistry P, Jackson S, Ackermann H-W, Kasatiya SS. 1995. *Bacillus cereus* phage typing as an epidemiological tool in outbreaks of food poisoning. J Clin Microbiol 33:636–640.
45. Donelli G, Guglielmi F, Paoletti L. 1972. Structure and physico-chemical properties of bacteriophage G. I. Arrangement of protein subunits and contraction process of tail sheath. J Mol Biol 71:113–125.
46. Grose JH, Casjens SR. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family *Enterobacteriaceae*. Virology 468-470C:421–443.
47. Dei R. 1989. Observations on phage-typing of *Clostridium difficile*: preliminary evaluation of a phage panel. Eur J Epidemiol 5:351–354. https://doi.org/10.1007/BF00144837.
48. Mahony DE, Clow J, Atkinson L, Vakharia N, Schlech WF. 1991. Development and application of a multiple typing system for *Clostridium difficile*. Appl Environ Microbiol 57:1873–1879.
49. Sekulovic O, Ospina Bedoya M, Fivian-Hughes AS, Fairweather NF, Fortier L-C. 2015. The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. Mol Microbiol 98:329–342. https://doi.org/10.1111/mmi.13121.
50. Hargreaves KR, Flores CO, Lawley TD, Clokie MRJ. 2014. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. mBio 5:e01045-13. https://doi.org/10.1128/mBio.01045-13.
51. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. Nat Commun 1:147. https://doi.org/10.1038/ncomms1146.
52. Haeusser DP, Schwartz RL, Smith AM, Oates ME, Levin PA. 2004. EzrA prevents aberrant cell division by modulating assembly of the cytoskeletal protein FtsZ. Mol Microbiol 52:801–814. https://doi.org/10.1111/j.1365-2958.2004.04016.x.
53. Paredes-Sabja D, Shen A, Sorg JA. 2014. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol 22:406–416. https://doi.org/10.1016/j.tim.2014.04.003.
54. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJJ, van der Oost J. 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell 46:595–605. https://doi.org/10.1016/j.molcel.2012.03.018.
55. Seed KD, Lazinski DW, Calderwood SB, Camilli A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature 494:489–491. https://doi.org/10.1038/nature11927.
56. Zheng Z, Bao M, Wu F, Chen J, Deng X. 2016. Predominance of single prophage carrying a CRISPR/cas System in "*Candidatus* Liberibacter asiaticus" strains in southern China. PLoS One 11:e0146422. https://doi.org/10.1371/journal.pone.0146422.
57. Sekulovic O, Fortier L-C. 2016. Characterization of functional prophages in *Clostridium difficile*. Methods Mol Biol 1476:143–165. https://doi.org/10.1007/978-1-4939-6361-4_11.
58. Sirard S, Valiquette L, Fortier L-C. 2011. Lack of association between clinical outcome of *Clostridium difficile* infections, strain type, and virulence-associated phenotypes. J Clin Microbiol 49:4040–4046. https://doi.org/10.1128/JCM.05053-11.
59. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
60. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C, Médigue C. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. Database (Oxford) 2009:bap021–bap021. https://doi.org/10.1093/database/bap021.
61. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM, Jeffery KJM, Jolley KA, Kirton R, Peto TE, Rees G, Stoesser N, Vaughan A, Walker AS, Young BC, Wilcox M, Dingle KE. 2010. Multilocus sequence typing of *Clostridium difficile*. J Clin Microbiol 48:770–778. https://doi.org/10.1128/JCM.01796-09.
62. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12:395. https://doi.org/10.1186/1471-2105-12-395.